
Review of Refugee Mental Health Assessment: Best Practices and Recommendations

Graham R. Davidson

University of the Sunshine Coast, Australia

Kate E. Murray

San Diego State University, USA

Robert D. Schweitzer

Queensland University of Technology, Australia

This article focuses on mental health assessment of refugees in clinical, educational and administrative-legal settings in order to synthesise research and practice designed to enhance and promote further development of culturally appropriate clinical assessment services during the refugee resettlement process. It specifically surveys research published over the last 25 years into the development, reliability measurement and validity testing of assessment instruments, which have been used with children, adolescents and adults from refugee backgrounds, prior to or following their arrival in a resettlement country, to determine whether the instruments meet established crosscultural standards of conceptual, functional, linguistic, technical and normative equivalence. The findings suggest that, although attempts have been made to develop internally reliable, appropriately normed tests for use with refugees from diverse cultural and linguistic backgrounds, matters of conceptual and linguistic equivalence and test-retest reliability are often overlooked. Implications of these oversights for underreporting refugees' mental health needs are considered. Efforts should also be directed towards development of culturally comparable, valid and reliable measures of refugee children's mental health and of refugee children's and adults' psychoeducational, neuropsychological and applied memory capabilities.

Keywords: assessment, mental health, cognitive, educational, refugees

Assessment is a cornerstone activity of psychological practice. Notwithstanding, and in spite of the recent publication of a number of useful handbooks on cross-cultural assessment (e.g., Suzuki & Ponterotto, 2008; Suzuki, Ponterotto, & Meller, 2001; Wilson & Tang, 2007), development and adaptation of measures that adequately assess the mental health and wellbeing status of culturally and linguistically diverse populations have continued to offer challenges for psychologists. Assessment of individuals from refugee backgrounds, in particular, provides additional challenges. In addition to the generally accepted challenges of establishing cross-cultural and measurement equivalence, which will be reviewed briefly below, refugee populations in transit or in the early stages of resettlement are often socially, cul-

turally and linguistically diverse groupings, creating the potential for cultural and linguistic barriers to have an undue influence on psychological test performance. They may be from cultures or linguistic groupings for which suitable, culturally and linguistically sensitive assessment instruments or established assessment norms are unavailable. They may be from nonliterate cultures. Research documenting the prevalence of mental ill-health in refugee populations also suggests they may be experiencing complex psychopathology and adjustment syndromes resulting from trauma and forced migration stress (Courtois, 2004; Davidson, Murray, & Schweitzer, 2008). They may be suffering from malnutrition and illness or have had some form of head trauma that result in short-term or long-term cognitive deficits (Weinstein,

Address for correspondence: Emeritus Professor G.R. Davidson, University of the Sunshine Coast, School of Social Sciences, Maroochydore DC, Sippy Downs, Qld 4558, Australia. E-mail: gdavidso@usc.edu.au

Fucetola, & Mollica, 2001). They may be suspicious of any form of government services and, as a consequence, reluctant to participate in psychological assessments or mental health interventions. Their psychological needs assessment may be broader than mental health *per se* and include psychoeducational assessment, neuropsychological testing, and cognitive assessment, including measures of memory recall within administrative-legal settings in which asylum claims are processed — keeping in mind research findings that suggest close associations between mental health and wellness on the one hand and educational adjustment, resettlement processing and a complex range of systemic and host community responses on the other. In summary, people and groups of refugee origins are subject to a wide range of ‘push’ factors, traumatic experiences and systemic mistreatment that are uncharacteristic of the migrant experience in general (Davidson, Murray, & Schweitzer, 2008). Hence, the need to focus specifically on their psychological assessment needs.

Valid and reliable assessment of refugees’ mental health and wellbeing should not be considered as an end in itself. Experiences of clients from refugee backgrounds, or of clients who have been internally displaced within their own national borders by war or by political, cultural or religious persecution, quite frequently involve direct exposure to traumatic events, including torture, loss of family and loved ones and witnessing violence and murder. Exposure to these types of trauma, in turn, has been linked with other lost opportunities and negative outcomes such as substance abuse, physical health problems, educational underperformance, illiteracy and lower incomes (Odejide, 2006; Robertson, Halcon, Savik et al., 2006). Inability to work, complex psychiatric disorders and educational underattainment, in turn, are reliable indicators of poverty. Valid and reliable mental health assessment methods and tools, therefore, are important for providing suitably targeted, culturally responsive mental healthcare in efforts not only to enhance mental health but also to enhance social, educational and occupational inclusion.

Furthermore, prevalence studies with clients from refugee backgrounds have resulted in inconsistencies in rates of reported PTSD, anxiety, depression and other disorders, which not only reflect differences in the trauma experience within specific cohorts but also possibly result from researchers and practitioners using different measures of psychological distress and wellbeing, applying different diagnostic cut-offs even when the same measure is used, being ignorant of culture-specific symptoms of distress, and not having an appreciation of other cultural or cohort-specific factors associated with distress (Davidson, Murray, & Schweitzer, 2008). Better knowledge of the cultural and linguistic sensitivity of a range of suitable assessment instruments, along with culturally appropriate norms for those instruments, may

provide more accurate information on prevalence rates and subsequent benchmarks for implementation and evaluation of interventions.

Previous reviews suggest that development or adaptation of clinical assessment instruments specifically for use with refugees is limited (Gagnon, Tuck, & Barkun, 2004; Hollifield et al., 2002). In the first of these large-scale reviews, Hollifield et al. (2002) reviewed literature on measures of refugee trauma and health status contained in six large online databases. Publications were evaluated according to the information they provided on their stated purpose, definitional clarity of the construct(s) being measured, research design, development process, reliability and validity. The 183 publications selected for detailed analysis on those criteria produced 12 measures that were designed specifically for assessment of trauma and health in refugee populations. None met all five evaluation criteria; three instruments (the Harvard Trauma Questionnaire [HTQ], Vietnamese Depression Scale [VDS] and the Adapted Hopkins Symptom Checklist–Depression [AHDSC; Bolton, 2001b]) met four of the criteria; and five instruments met only one criterion. The review also identified eight measures that were developed initially for general use but were subsequently used with refugee populations. Two of the eight measures (the Hopkins Symptom Checklist [HSCL] and the Beck Depression Inventory [BDI]) met all five evaluation criteria; and the remaining six measures (Impact of Events Scale [IES], Symptom Checklist–90 [SCL-90], Health Opinion Survey [HOS], Cornell Medical Index [CMI], Posttraumatic Symptom Scale–10 [PTSS-10] and the Norbeck Social Support Questionnaire) met four of the five criteria. Hollifield et al. (2002) identified as issues in refugee assessment the need for clear, theory-driven models of trauma and health assessment and the desirability of combining qualitative and quantitative research when developing instruments, in order to improve their linguistic and cultural comprehension.

Gagnon et al. (2004) reviewed over 10,000 articles employing measures of health among women from refugee and other migrant backgrounds. The studies they reviewed were not necessarily concerned with the development of assessment instruments for those purposes but were dealing primarily with the application and utilisation of women’s health measures. Gagnon et al. deemed the selected studies and measures to be relevant and ‘high quality’ if they met the following criteria: the measure should provide a comprehensive assessment of women’s mental health status and related health issues; women should constitute at least 50% of the sample; there should be evidence of both *emic* and *etic* analysis so that measures have both local and comparative diagnostic value; and there should be evidence that established translation procedures were used in the measure’s preparation. Only five out of the 47 measures that were analysed met those

criteria, the list of high quality measures being: the Harvard Trauma Questionnaire, the General Health Questionnaire-28, the Hopkins Symptom Checklist-25, the Vietnamese Depression Scale and the Beck Depression Inventory. Gagnon et al. (2004) suggested that the existing refugee literature and measures are underutilised and that published research underreports reliability and validity of those measures.

A practitioner-focused review of refugee mental health assessment and treatment was provided by Ehntholt and Yule (2006). However, their review did not offer specific criteria for judging the scientific quality of assessment instruments. In addition to recommending semistructured and structured clinical interviews as appropriate assessment approaches, the review also suggested the following instruments be used: the War Trauma Questionnaire (WTQ), Impact of Events Scale-Revised for Children (IES-R), Birlerson Depression Self-rating Scale (BSRS), Revised Children's Manifest Anxiety Scale (CMAS-R) and the Strengths and Difficulties Questionnaire (SDQ). However, Ehntholt and Yule (2006, p. 1202) concluded that 'Development of new measures designed specifically for young refugees or at the very least studies that investigate the reliability and validity of using existing measures with this population are urgently required'.

The aim of the current review was not to provide comprehensive details on the development and validation of multicultural assessment instruments; nor did it aim to critique the ethical and professional applications of instruments developed for general crosscultural use. Readers are referred to other more detailed resources for of those issues (e.g., Dana, 2005, 2007; Davidson, 1997; Suzuki & Ponterotto, 2008; Suzuki, Ponterotto, & Meller, 2001). There are also useful general guides for assessing posttraumatic symptoms in refugees and asylum seekers (e.g. Drozdek & Wilson, 2007; Okawa, 2008). The specific aims, instead, were to (a) survey assessment instruments that have been developed or adapted specifically for use with refugee clients and tested thoroughly in those assessment contexts and (b) identify areas of assessment for which there is an absence of tried and tested assessment instruments for use with clients from refugee backgrounds. The purposes of the review were to assess progress since the two earlier large-scale reviews on developing or adapting mental health measures specifically for use with these clients and, in doing so, to highlight specific, key issues related to their development and standardisation, with a view to assisting future prevalence research and measurement of treatment outcomes.

Method

Selection Criteria

Assessment measures were identified by interrogating PsycINFO and PUBMED to locate studies with refugees

and similarly displaced persons, which were published in English in the last 25 years. The database searches specified the following terms or variations thereon: refugee, asylum seeker, internally displaced, assessment, mental health, educational, cognitive and neuropsychological. Research dissertations were excluded from the search.

Studies were selected for tabulation if their specific purpose was the development, adaptation, or examination of assessment instruments or calculation of their statistical validity and/or reliability with refugee clients from specific cultural and linguistic backgrounds. That is, studies were selected for analysis only if they specifically involved development of a new instrument for use with refugee clients or adaptation or restandardisation of an existing instrument for that purpose. The review, therefore, differed from the earlier reviews of Hollifield et al. (2002) and Gagnon et al. (2004) because it deliberately excluded studies (a) that reported on the analysis of measures designed generally for crosscultural use and (b) where the reported analysis of the properties of a measure was incidental to the main purpose of the study.

There is good reason to limit the review according to these parameters. First, a case has been made that assessment of clients from refugee backgrounds may be far more complex and demanding than assessing culturally and linguistically different clients or voluntary migrants in general. Second, incidental reporting of statistical properties of assessment measures used for other research or professional purposes frequently relies on statistical values previously reported in the literature rather than recalculating the values for the sample under investigation or may overstate, or simply not report, the values if they cast doubt on the study's core aims and objectives. In summary, the review we conducted looked specifically at psychometric and other related research specifically into the development or adaptation of measures of refugees' mental health and wellbeing.

Evaluation Criteria

Although the database searches concentrated specifically on published research involving development, adaptation or restandardisation of measures used with persons from refugee backgrounds in either the internal displacement, flight, or resettlement stages, nevertheless research into, and applications of, these measures normally involve working with a range of culturally and linguistically different clients. Therefore, normally accepted standards that allow for crosscultural comparison apply. There is a range of criteria by which crosscultural comparability of alternate forms of a measure may be assessed, including construct (or conceptual) equivalence, functional equivalence, linguistic equivalence, contextual equivalence and normative equivalence (see e.g., Berry, 1969; Okawa, 2008; van de Vijver & Poortinga, 2005). Put succinctly, culturally comparable versions of measures have the following

properties. They demonstrate conceptual equivalence in that the constructs they assess will have the same meaning in test takers' culture of origin and in the culture for which the measure was originally developed. They demonstrate functional equivalence in that assessment items measure how a construct is manifest in test takers' culture of origin and in the culture for which the measure was originally developed. They demonstrate linguistic equivalence in that test takers and assessors share a common understanding of items' linguistic meanings, that is, the original and translated language versions of a measure should have the same linguistic meaning. Finally, but not exhaustively, comparable versions should be contextually equivalent in the sense that the context, method and circumstances of their administration are suitable and appropriate for test takers.

The extent to which alternate versions of a measure are conceptually, functionally and contextually equivalent or, alternatively, the development of culturally comparable versions, is best achieved by researchers undertaking in-depth ethnographic analysis of the constructs under investigation within the culture or linguistic group in which the alternate version will be administered. This level of analysis may also be achieved successfully by having either local indigenous or cultural 'experts' examine and offer advice on the construct being measured, how it may be manifested in the second cultural or linguistic context and how it would be appropriate to elicit information from individuals within the second cultural or linguistic context. An important element of estimating cultural comparability is the understanding of what is contextually appropriate investigative behaviour, specifically, is asking questions the most appropriate way to obtain the information? In what circumstances is it appropriate to ask questions? What factors determine who may ask questions of whom? Might information best be obtained using an oral or a written format? How familiar are informants with formal questionnaires or interviews? Such questions about data gathering once more are best addressed by an in-depth ethnographic analysis of the cultural context of planned inquiries.

Linguistic equivalence is best estimated by using one or, preferably, a combination of tried and tested techniques, including back-translation, committee evaluation and bilingual knowledge and/or performance testing (Brislin, Lonner, & Thorndike, 1973; Cha, Kim, & Erlen, 2007; Mallinckrodt & Chia-Chih, 2004; Sechrest, Fay, & Zaida, 1972; Sinaiko & Brislin, 1973).

Researchers should attempt to estimate the extent of normative equivalence of assessment instruments designed for use crossculturally. Alternate versions of an instrument that are normatively equivalent will produce similar mean, variance and standard error estimates. If normative equivalence cannot be established, the alternate version should be re-normed for the population for

which it is intended. The statistical properties of alternate versions of a measure also need to be explored and reported (Ponterotto, 1996). In addition to the measures having face validity, the questions should be comprehensible and appropriate for those being assessed and, in the absence of longitudinal information that allows for the calculation of their predictive value, the versions should have both concurrent and discriminant validity in that they should accord with assessments obtained by other measures of related constructs and diverge from assessments of unrelated constructs. Finally, instruments should be reliable as measured by various methods for calculating internal consistency or by the split-half method; and, all other things being equal, results should be stable across time and context, as measured by the retest method. It is customary, if not necessary, for test developers and users to report on these statistical properties of the measures being used, including the properties of alternate versions of those measures.

Consequently, in our review of relevant, reported assessment instruments, attention was given to whether authors (a) undertook an in-depth intracultural analysis of the measures being used, (b) provided detailed information on the use of accepted translation techniques, (c) reported on internal consistency and retest reliability and (d) provided evidence of concurrent or discriminant validity.

Results and Discussion

Diagnostic and Clinical Assessment

The review identified only 17 studies in total that met the selection criteria. Eleven (64.7%) of the studies reported on the development, adaptation or examination of measures designed for use with adults, while five reported on measures designed for use with adolescents or young adults. Only one study reported on a measure adapted for use specifically with young children from refugee backgrounds. Analysis of the studies is presented in Table 1. There is a demonstrable need for further development or adaptation of scales designed specifically to assess young children's mental health and wellbeing following their experiences of war, internal displacement or flight and resettlement.

In six cases (35.3%) evidence was presented to indicate that the researchers had attempted to establish conceptual (or construct) equivalence of the measure for the population being assessed. This evidence was in the form of detailed reporting of an ethnographic analysis of the concepts being measured, or interviews with knowledgeable local elders or experts, or focus groups. In eight cases (47%) there was evidence that the researchers had undertaken thorough back-translation of the measure(s) into the first language of those being assessed by using the back-translation method by itself or in combination with a committee or bilingual assessment approach (Brislin, Lonner, & Thorndike, 1973). Although two

Table 1

Assessment Instruments Developed or Adapted for Specific Use with Refugee Populations

Author	Scale	Culture	Evidence of Construct Equivalence	Evidence of Linguistic Equivalence	Internal Consistency	Test-Retest Reliability	Concurrent Validity
Adult Scales							
Bolton (2001b)	Adapted Hopkins Symptom Checklist-Depression (AHDSCD)	Rwandan war victims Study 1 $n = 93$; Study 2 $n = 368$	Ethnographic analysis	B-T + C	$\alpha = 0.87$	$r = 0.67$ ($n = 37$)	Community assessment $\Phi = 0.38$ ($n = 93$)
Charney & Keane (2007)	Clinician Administered PTSD Scale (CAPS)	Bosnian refugees $n = 115$	—	B-T + C	$\alpha = 0.92$	—	Correlated with Beck Depression Inventory-II (BDI-II), Structured Clinical Inventory (SCID), General Health Questionnaire (GHQ) $p < .01$; SISOW $p < .05$
Fawzi et al. (1997)	Harvard Trauma Questionnaire (HTQ)	Vietnamese refugees in the USA $n = 74$	—	—	$\alpha = 0.78-0.92$	—	Severity and events correlated $p < .0001$
Hinton et al. (2006)	Anxiety Sensitivity Index (ASI) Augmented ASI (AASI)	Cambodian refugees in the USA $n = 208$	Ethnographic analysis	B-T + BTA-K	AASI $\alpha = 0.36-0.91$	ASI $r = 0.89$; AASI $r = 0.83$	—
Hoffmann et al. (2005)	Short Form Health Survey (SF-12)	Russian-speaking refugees $n = 59$	Ethnographic interview	B-T	$\alpha = 0.66-0.70$	—	Clinic < Community sample $p < .01$
Hollifield et al. (2005)	Comprehensive Trauma Inventory (CTI-164)	Kurdish & Vietnamese refugees in the USA $n = 67$	Interviews and focus groups	—	Subscale—total score $r = 0.52-0.99$	—	Sig differences ($p < .05$) between entry criteria categories on Trauma Experiences Questionnaire (TEQ)
Lavik et al. (1999)	Hopkins Symptom Checklist (HSC-25)	English-, Arabian-, Persian-, Turkish-, Vietnamese-, and Norwegian-speaking outpatient clinic refugees in Norway $n = 231$	—	—	—	—	Global Assessment Scale (GAS) $p < .0001$ Brief Psychiatric Rating Scale (BPRS) Total $p < .000$
Malekzai et al. (1996)	Clinician Administered PTSD Scale (CAPS-1)	Afghan (Pushto & Farsi) refugees in the USA $n = 30$	—	C	$\alpha = 0.76-0.88$	*	—
Miller et al. (2006)	Afghan Symptom Checklist (ASCL)	Adults from Kabul districts $n = 324$	Analysis of local trauma narratives	B-T + C	$\alpha = 0.74-0.93$	—	Afghan War Experiences Scale (AWES) $r = 0.70$
Mollica et al. (1992)	Harvard Trauma Questionnaire (HTQ)	Cambodian, Laotian, Vietnamese refugees $n = 91$	—	B-T + C (Khmer, Lao, Vietnamese)	$\alpha = 0.90$ (trauma events) $\alpha = 0.96$ (trauma symptoms)	$r = 0.89-0.92$	DSM III-R PTSD vs. Non-PTSD $p < .0001$. Better predictor of DSM III-R diagnosis than HSC-25

(continued over)

Table 1 (continued)

Assessment Instruments Developed or Adapted for Specific Use with Refugee Populations

Author	Scale	Culture	Evidence of Construct Equivalence	Evidence of Linguistic Equivalence	Internal Consistency	Test-Retest Reliability	Concurrent Validity
Renner et al. (2006)	Clinician Administered PTSD Scale (CAPS-1) Hopkins Symptom Checklist (HSCL-25) Harvard Trauma Questionnaire (HTQ) Impact of Events Scale—Revised (IES-R) Bradford Somatic Inventory (BSI) Social Adaptation Self-Evaluation Scale (SASS)	Chechen, Afghan, West African <i>n</i> = 150	Expert interviews	B-T	HSCL-25 α = 0.91–0.96; HTQ α = 0.87–0.91; IES-R α = 0.91–0.96; BSI α = 0.95–0.97; CAPS-1 α = 0.90–0.91; SASS α = 0.48–0.74	—	CAPS-1 & IES-R predicted clinical vs. non-clinical group membership based on clinical interview <i>p</i> < .01 for 3 ethnic groups
Adolescent and Young Adult Scales							
Bean, Derluyn et al. (2006)	Reactions of Adolescents to Traumatic Stress (RATS)	Resident refugees in Belgian & Holland speaking 46 languages <i>n</i> = 1,859 refugees; <i>n</i> = 1,676 Dutch & Belgian comparisons	—	C	α = 0.81–0.93	<i>r</i> = 0.44–0.63 (<i>n</i> = 519)	Sig <i>r p</i> < .001 with HSCL-37 & Stressful Life Events (SLE)
Bean, Derluyn et al. (2007)	Hopkins Symptom Checklist-37 (HSCL-37)	Resident refugees in the Netherlands speaking 46 languages <i>n</i> = 2,214 refugees; <i>n</i> = 1,979 Dutch & Belgian comparisons	—	C	α = 0.84–0.95	—	Sig <i>r p</i> < .001 with RATS, SLE and Strengths & Difficulties Questionnaire (SDQ)
Bean, Mooijaart et al. (2006)	Child Behaviour Checklist (CBCL) for Guardians of Unaccompanied Minors	Guardians of resident refugee minors from 48 countries in the Netherlands <i>n</i> = 478 URM	—	Administered in Dutch	α = 0.94	—	Sig <i>r p</i> < .01 between URM, teacher, and guardian
Bean, Mooijaart et al. (2007)	Teacher Report Form (TRF) for Unaccompanied minors	Teachers of URM in the Netherlands <i>n</i> = 486 URM	—	Administered in Dutch	α = 0.55–0.95	—	Internalising score differences for MHC and no MHC groups <i>p</i> < .01
Sack et al. (1998)	Impact of Events Scale (IES)	Khmer refugees in the USA <i>n</i> = 180	—	—	α = 0.92	—	DSM III-R clinical > nonclinical group <i>p</i> < .001
Children's Scales							
Smith et al. (2003)	Impact of Events Scale (IES-13)	Bosnian children in war <i>n</i> = 2,976	—	B-T	α = 0.60–0.80	—	—

Key: B-T = Back-translation; C = Committee; BTA-K = Bilingual test administration—knowledge testing; * Item interrater reliability 0.86–1.00; URM = Unaccompanied refugee minor; MHC = Mental healthcare; SSOW = Semistructured Interview for Survivors of War

studies (Bean, Derluyn et al., 2006; Bean, Derluyn et al., 2007) contained a description of translation attempts that might approximate the committee approach, there was insufficient information to allow the reader to assess the stringency of its application. Two (11.8%) studies involved appropriate administration of measures to local guardians or teachers in the national language, although their conceptual equivalence for those many cultures from which young refugees came was not questioned (Bean, Mooijaart et al., 2006; Bean, Mooijaart et al., 2007). Therefore, judging by the criteria set for this review, five (29.4%) of the studies appeared to give insufficient attention both to issues of conceptual equivalence and to accepted, detailed translation techniques.

Conceptual and Linguistic Sensitivity

Just over a third of the selected studies reported on efforts to conduct an ethnographic analysis of the concepts being measured, or hold interviews with knowledgeable local elders or experts, or focus groups. Following a sequence of interviews and focus groups, Hollifield et al. (2005) developed the Comprehensive Trauma Inventory that includes 164 items assessing traumatic experiences among refugees. In their assessment, refugees experienced an average of 150 traumatic events, much higher than the numbers suggested by previous measures. Hollifield et al. (2005) attributed the differential outcomes to limitations in the design of alternative measures. Development and statistical analysis of the Comprehensive Trauma Inventory illustrates how measures that are sufficiently detailed to capture indigenous experiences of trauma may more accurately and comprehensively capture survivors' mental health status than do measures that have been developed for general use and evaluated previously with other populations.

Measures may also emphasise different phenomena that are more or less important in different cultures. For example, after analysing detailed trauma narratives obtained from Kabul residents, Miller et al. (2006) developed a symptom checklist using more culturally appropriate idioms of distress to capture symptoms of distress among displaced Afghans. This 22-item measure was developed from narrative examples of many common symptoms of distress existing in western measures (e.g., loss of appetite, crying) but also included seven indigenous items. Miller et al. maintained their research was premised on a social constructivist approach to clinical assessment, which does not repudiate the possible existence of universal types of responses to traumatic events, but 'caution[s] against assuming that psychiatric constructs identified in one culture are similarly meaningful in other cultural contexts' (Miller et al., 2006, p. 424). Using in-depth ethnographic analysis of interview data, Bolton (2001a) examined two local Rwandan culture-specific syndromes — a mental trauma and a grief syndrome — that shared some symptoms

with the DSM-IV PTSD and grief syndromes. Culture-specific symptoms of mental trauma and grief were subsequently added to the Hopkins Symptom Checklist–Depression, which was administered to survivors (Bolton, 2001b). A parallel analysis of depression and anxiety among northern Ugandan war survivors resulted similarly in an overlap between the DSM-IV diagnoses and related, local symptoms, as well as in the identification of a number of culture-specific symptoms and clusters (Bolton et al., 2007). Exploration and establishment of equivalence between original, western and local indigenous constructs and symptoms of mental health and wellness or, alternatively, application of local, indigenous constructions in assessing mental health is an important first step towards ensuring that the assessment instruments are tapping into clients' understandings of distress and wellness.

The attempt by Miller et al. (2006) to incorporate etic and emic expressions of psychological distress into their symptom checklist is congruent with the debate about the applicability and appropriateness of western measures for assessing clients from nonwestern backgrounds. This is evident in the debate surrounding the applicability and significance of PTSD as a diagnostic category with refugees (Bracken, Giller, & Summerfield, 1995; Kagee & Naidoo, 2004; Summerfield, 1999). Silove (2006) argued practitioners must recognise the difference between 'self-limiting trauma reactions and longer-lasting disabling psychopathology' (p. 15) and that the experience of PTSD may not be the primary concern among resettling refugees. Investigators increasingly are recognising that manifestation, meaning and significance of distress may be different across cultures and contexts. Therefore, effective assessment and treatment needs to identify patterns of distress among a specific population and explain how those symptoms are developed, understood and treated within the culture of origin (Hinton et al., 2005).

Other researchers recommend that psychologists need to look beyond the traditional focus on PTSD and the victimisation of refugees (Muecke, 1992) in order to examine the broader context and various dimensions that are a part of their lives (Brough, Gorman, Ramirez, & Westoby, 2003). Papadopoulos (2007) argued for the need to consider mental health outcomes from a broader sociopolitical, rather than a narrow individual symptom-focused, perspective. Therefore, when assessing the experiences of refugees, assessment must move beyond clinical measures of PTSD and pathology and examine the breadth of the human experience in resettlement, including the ways in which people make sense of, and grow through, their refugee experiences. Development or adaptation of actuarial tools that measure personal growth and change specifically with clients who have experienced armed conflict, displacement, flight and resettlement is recommended.

The need for culturally competent child intake assessments to address cultural identity, cultural explanations of presenting concerns, cultural factors related to the psychosocial environment and level of functioning, and cultural elements of the relationship between the individual and the clinician was emphasised by Ecklund and Johnson (2007). They also provided summary suggestions and culturally relevant probes to guide practitioners. Other writers have suggested detailed models that outline many of the predisplacement, postdisplacement, systemic and outcome variables to which clinicians should attend when assessing and treating clients from a refugee background (Davidson, Murray, & Schweitzer, 2008; Gagnon, Tuck, & Barkun, 2004). Given the narrow focus of current refugee clinical assessment tools, such models are useful in providing a comprehensive set of criteria for practitioners to address when initial assessments are conducted.

Almost half of the selected studies reported the use of recognised procedures for checking translations. Back-translation was used most frequently, either separately or in combination with bilingual committees, to resolve discrepancies after back-translation had been undertaken. Two notable exemplars that adopted a slightly different approach were the study by Hoffmann et al. (2005), which employed back-translation followed by administration of both alternate language versions to expert bilingual test takers who were asked to identify wording discrepancies, and the study by Malekzai et al. (1996), which employed a panel of five expert bilingual physicians living in the United States to evaluate and comment separately on the English and translated Pushto and Farsi versions of the CAPS-1. Nevertheless, two items still presented problems for the majority of Malekzai and colleagues' 30 respondents. As examples of decentring following back-translation, the study by Bolton (2001b) of the Adapted Hopkins Symptom Checklist–Depression (AHDSC) commented on the researchers' replacement of western terms with local terms when both terms were used interchangeably by the bilingual translators and committee members. Miller et al. (2006) mentioned the need for numerous changes required to ensure the English and Dari versions of the Afghan Symptom Checklist (ASCL) were linguistically equivalent. By way of contrast, the studies by Bean, Derluyn et al. (2006, 2007) respectively of the Reactions of Adolescents to Traumatic Stress (RATS) and Hopkins Symptom Checklist-37 (HSCL-37) in 47 different languages stated explicitly that back-translation was not used but that judgments about equivalence were left to translators and interviewers. These studies were not considered as having met the linguistic equivalence criterion.

Statistical Criteria

All 17 studies contained sufficient information on mean tendencies and variance to allow for the establishment of

sample-specific norms. Only one study (Lavik et al., 1999) did not report internal reliability (α) coefficients. Internal reliabilities for the remaining measures were uniformly good to very high, with α coefficients being generally higher for total scale scores than for subscale scores. The study by Renner et al. (2006) of the Clinician Administered PTSD Scale (CAPS-1), Hopkins Symptom Checklist (HSCL-25), Harvard Trauma Questionnaire (HTQ), Impact of Events Scale–Revised (IES-R), Bradford Somatic Inventory (BSI), Social Adaptation Self-Evaluation Scale (SASS) demonstrated that internal reliability may differ according to the linguistic version used, but those linguistic factors appeared to have less influence on reliability than the whole scale vs. subscale distinction.

Only four (23.5%) of the seventeen studies reported test–retest reliability, which varied from moderate to high depending on the study. Bean, Derluyn et al. (2006) reported slight differences in test–retest reliability for total scale and subscale scores on the Reactions of Adolescents to Traumatic Stress (RATS). The large majority (82%) of studies reported on concurrent, criterion or discriminant validity in one form or another.

Clinical vs. Actuarial Assessment

Ehnholt and Yule (2006) recommended semistructured and structured clinical interviews be used as appropriate in conjunction with the actuarial measures they surveyed. This raises the question about the advantages and disadvantages of using these broad types of assessment with clients from refugee backgrounds. There is little discussion of their relative merits in the literature. However, after reviewing literature on the use of survey approaches to assess the impacts of trauma on victims of torture, Willis and Gonzalez (1998, p. 283) concluded that 'if properly conducted, the survey approach represents an effective method' provided that: respondents are informed in an appropriate fashion about the purpose and rationale of the survey; they can see benefit in the survey either for themselves or their community at large; data are used carefully and judiciously for diagnostic, therapeutic or administrative purposes; and researchers are sensitive to cultural factors that may influence respondents' performance.

By way of comparison, according to Kirmayer (2003), full comprehension of the experiences of refugees and other survivors of war-related atrocities may be difficult to achieve within a clinical assessment paradigm. The limitations may be due mutually to clinicians and their clients. On the one hand, clinicians may lack the required cultural and experiential knowledge — Kirmayer uses the term 'imagination' — to reconcile survivors' stories of atrocities and other human rights violations, displacement and relocation with their own understandings of what is possible and understandable within people's life journeys. As a consequence, clini-

cians may interpret survivors' experiences within a familiar psychopathological paradigm wherein, for example, the unfamiliar is regarded as disorder, delusion, paranoia or confabulation. Survivors in the role of client may lack the skills and knowledge of the host culture and the clinician to be able to construct a coherent, comprehensible account of their exceptional experiences. They may not understand what they should disclose, they may not be able adequately to find words or suitable story structures to convey the chaotic nature of their survival and they may be initially incapable of comprehending that there is possibly a deeper level of meaning to their experience. The possible consequences of this disjunction between the exceptional life narratives of survivors and the clinical and administrative narratives of health practitioners and officials are that survivors will receive inappropriate professional treatment and unfair adjudications of claims for refuge. The solutions to these narrative misunderstandings may involve giving survivors clear and detailed explanations of the purpose and context in which their story is sought, including the diagnostic, therapeutic or administrative purposes for which information will be used, combining clinical with actuarial assessments to reach a considered judgment about survivors' state of mind and, very importantly, training clinicians and administrative officials to comprehend the storied nature of memories — a matter to be considered below.

Educational and Neuropsychological Assessments

The review failed to uncover any psychoeducational assessment tools developed or adapted specifically for use with school students from refugee backgrounds. This is despite the fact that there are concerns about how such students are assessed and assisted when they enter school following resettlement (Fraine & McDade, 2009). Measures of educational aptitude and achievement need to recognise the unique experiences, educational history and emotional state of newly arrived refugee students (Dao, 1991; Rousseau, Drapeau, & Corin, 1996). There are various ways in which psychologists and teachers could recognise and respond to the unique experiences and backgrounds of children and families from a refugee background. For example, children and families may have experienced a significant disruption in education during their period of flight or in refugee camp settings. Some refugees come from highly aural cultures, where less emphasis is placed on written knowledge and understanding, which may have influenced their previous schooling (Burgoyne & Hull, 2007). Dao (1991) highlights five issues that need to be addressed when assessing refugee students in school settings, including linguistic and cultural background, acculturation problems, literacy and basic skill levels, problem-solving skills and emotional difficulties. Fraine and McDade (2009) have also proposed a comprehensive five-step program for orientation, assess-

ment and support of students from refugee backgrounds when they enter a new school. Despite the value obtained from these orderly entry and support frameworks, they do not appear to resolve the frequently acknowledged problem of students being age graded in schools, rather than being graded in accordance with their educational and language abilities at point of entry. Nor do they appear to solve the conundrum of how a relatively inflexible age grading system manages older young people whose educational and language abilities on entry are incommensurate with their age grade.

Undertaking assessments of intellectual functioning in children from refugee backgrounds is a complex task that involves an appreciation of both the crosscultural cognitive ability testing tradition and the refugee experience. There are few options that meet the needs of practitioners, crosscultural comparability principles and gold psychometric standards. The Victorian Department of Education and Early Childhood Environment (2009) has developed guidelines for assessment of children from refugee and other non-English-speaking backgrounds that recommend the use of five tests: the Wechsler Nonverbal Scale of Ability (WNV), the Universal Nonverbal Intelligence Test (UNIT), Raven's Progressive Matrices (RPM) and the Naglieri Nonverbal Ability Test-Individual (NNAT-Individual). The guidelines indicate that these psychometric measures of cognitive ability should be used in conjunction with an assessment of the student's adaptive behaviour and history, including the family's culture and refugee experience, to arrive at an adequate assessment of students' psychoeducational needs. The guidelines do not offer direction on the norms against which these students' performance should be compared.

Our review of assessment of people from refugee backgrounds did not find any published studies reporting on the use of these four recommended gold standard, culture-reducing, nonverbal tests of cognitive ability with persons from refugee backgrounds. (The only reported work was by Mulder, 1992.) In the absence of such research, it is difficult to determine whether norms for test takers from refugee backgrounds should be similar to the norms for nonrefugees from similar cultural backgrounds or to majority culture norms. Therefore, it is difficult to determine whether individual deviations from the published norms are indicative of a cultural or educational difference or of the cognitive impacts of the refugee experience. Furthermore, in the absence of published research that reports on the reliability and validity of these measures of the ability of students from refugee backgrounds who may have complex trauma or other mental health symptoms, results on these tests for students from such backgrounds must be treated with extreme caution.

Psychological disorders such as PTSD may confound performance on educational assessments. For instance,

symptoms of PTSD have been linked to poorer cognitive functioning among adult refugees (Kivling-Boden & Sundbom, 2003). Ability to concentrate and intrusive thoughts may interfere with educational test performance. Physical health (e.g., brain injury, malnutrition) and psychological distress (e.g., depression, PTSD) may precipitate physiological changes in the brain. Weinstein, Fucetola, and Mollica (2001) highlighted the possibility of fixed neural loss and neuropsychological abnormalities that are triggered by traumatic events and further influenced by ongoing stress. They also highlighted the need for neuropsychologists to assess the impact of traumatic brain injury and experiences of torture and trauma using culturally sensitive measures. The complex neuropsychological, cultural and experiential factors involved in educational and psychological assessment are not clearly understood. Further research needs to be undertaken to develop and evaluate suitable educational and neuropsychological instruments for assessing the educational potential and neuropsychological functioning of young people from refugee backgrounds.

Autobiographical Recall

Assessment of memory and recall is critical to the administrative processing of refugee claims. The aforementioned issues are particularly salient in assessing the impact that refugee clients' cognitive impairment might have on this process. Only a few studies have examined the complexities involved in assessment when evaluating refugee claims. One study examined inconsistencies in refugee claimants' reports across interviews to determine whether those errors should serve as indicators of lack of credibility (Herlihy, Scragg, & Turner, 2002). The researchers found that discrepancies often occurred when individuals were interviewed on multiple occasions. The number of discrepancies for individuals with high levels of posttraumatic stress increased with increasing time between the interviews. In addition, they found greater divergence in the details refugee claimants rated as being peripheral to their account than in the details they rated as central to their claims.

One study of the Canadian system for reviewing refugee claims recognised the process as 'one of the most complex adjudication functions in industrialized societies' (Rousseau, Crepeau, Foxen, & Houle, 2002, p. 43). In that study, the researchers examined 40 difficult cases that were referred to the research team. In reviewing already decided cases, they uncovered a number of issues including: 'difficulties in evaluating evidence, assessing credibility, and conducting hearings; problems in coping with vicarious traumatisation and uncontrolled emotional reactions; poor knowledge of the political context, false representations of war, and cultural misunderstandings or insensitivity' (Rousseau et al., 2002, p. 43). Kirmayer (2003) re-analysed some of the examples of Canadian Immigration and Refugee Board hearings

Rousseau et al. (2002) provided where, under procedures for tendering information and subsequently cross-examination, doubt was cast on refugees' accounts of their experiences. According to Kirmayer, board hearings seek evidence in the form of a 'fixed, reliable, and reproducible account of historically verifiable events' (p. 170) based on information provided in claimants' Personal Information Forms relating to their 'language, nationality, ethnic group, religion, education, military service, work history, travel, prior claims for refugee status and any crimes or offences' (p. 173). Kirmayer argued that spontaneous narratives elicited from refugees for clinical or administrative purposes may appear to be structurally complicated, fragmentary, contradictory and formulaic, because they are culturally and personally complex, have been compacted temporally and frequently contain psychologically distressing material for which words are hard to find or about which claimants have been prevented previously from speaking.

The issues pertaining to the complexity of asylum claims need to be addressed because they are crucially significant for asylum policies and processing worldwide. Appropriate 'cognitive interviewing' instruments need to be developed and training in their use for administrative personnel is required if reviews of refugees' asylum applications are to become psychologically defensible and procedurally fair.

Summary

This review provides support for previous findings that the number of valid and reliable measures of refugees' mental health and related wellbeing is limited (Ehnholt & Yule, 2006), especially where assessment of young people from refugee backgrounds is concerned. Only one study involving an adaptation of the Adapted Hopkins Symptom Checklist–Depression (AHDSC) for use with Rwandan survivors of war met all five evaluation criteria in the current review. The following adult scales (see Table 1) met four of the five evaluation criteria, which included either ethnographic analysis to establish construct equivalence and/or thorough translation to establish linguistic equivalence: the Anxiety Sensitivity Index (ASI) and the Augmented ASI (AASI) with Cambodian refugees; the Short Form Health Survey (SF-12) for Russian-speaking refugees; the Comprehensive Trauma Inventory (CTI-164) for Kurdish and Vietnamese refugees; Afghan Symptom Checklist (ASCL) for Kabul-based adults; the Harvard Trauma Questionnaire (HTQ) for Cambodian, Laotian and Vietnamese refugees; and the Clinician Administered PTSD Scale (CAPS-1), Hopkins Symptom Checklist (HSCL-25), Harvard Trauma Questionnaire (HTQ), Impact of Events Scale–Revised (IES-R), Bradford Somatic Inventory (BSI) and Social Adaptation Self-Evaluation Scale (SASS) for Chechen, Afghan, and West African refugees. More research needs to be undertaken

for the purpose of developing or adapting assessment instruments for use with adolescents and children from refugee backgrounds. Further work is also required to develop culturally sensitive and valid measures of refugees' cognitive functioning, educational attainment, and psychoeducational adjustment. Greater emphasis should be given to neuropsychological assessment of refugees who have experienced torture, trauma or physical deprivation. Development of new assessment measures or the adaptation of existing measures should include provision for establishing construct and linguistic comparability across refugee populations, reporting on retest reliability and estimating reliability and validity for equivalent cultural and linguistic versions of measures.

Services for clients from refugee backgrounds, first and foremost, should be designed to assist refugees to 'develop a sense of stability, safety and trust, as well as to regain a sense of control over their lives' (Ehnholt & Yule, 2006, p. 1202). Use of valid and reliable methods for assessing their mental health status, cognitive capacity, neuropsychological functioning, psycholinguistic ability and psychoeducational needs is an important first step in meeting that objective. Supplementation of actuarial assessment with semistructured clinical assessments that elicit refugee clients' personal and cultural narratives of their experiences and distress is encouraged; and ethnographic, linguistic and statistical analysis of measures of subsequent personal growth and change specifically with clients who have experienced armed conflict, displacement, flight and resettlement is recommended. Psychologists also need to find ways of offering expert input on methods for evaluating claims for refugee status.

Limitations

It is difficult with such large-scale reviews to ensure that the outcome is a perfectly detailed statement of the literature. The results will be limited in the first instance by the choice of electronic databases and subsequently by the search terms entered. Therefore, readers should bear in mind that searches were limited to articles in English published since 1984 and abstracted in PsycINFO and PUBMED. Second, making the judgment that a published article is primarily about the development, adaptation or standardisation of an assessment instrument, rather than about the use of the instrument for diagnostic or treatment purposes, involves a certain level of subjectivity. Therefore, studies that have reported on diagnostic and therapeutic applications of assessment instruments may offer further information on the cultural sensitivity and the reliability and validity of those instruments in other cultural and linguistic contexts. Nevertheless, making a judgment based on purpose of the reported research is important because it is possible that studies designed specifically to investigate the cultural, linguistic, and statistical properties of an

instrument are more likely to do so thoroughly than diagnostic or therapeutic research for which analysis of those properties is a secondary aim. Finally, it might be suggested that the net should have been cast more widely to include all crosscultural studies of mental health, neuropsychological, cognitive and psychoeducational assessment rather than limiting the search to research involving refugees, asylum seekers and internally displaced victims of conflict. We acknowledge that crosscultural studies of psychological assessment have the potential to inform assessment of persons from refugee backgrounds. Nevertheless, assessment measures developed generally for use crossculturally need to be evaluated for use specifically with refugee populations, keeping in mind the complexity of 'push' factors, traumatic distress, systemic mistreatment, flight and resettlement experience and adjustment challenges that are part and parcel of the refugee experience.

Conclusion

Despite the limitations to this review, its findings are not inimical to, and are possibly more encouraging than, the findings of the other large-scale reviews by Gagnon, Tuck, and Barkun (2004) and Hollifield et al. (2002), albeit that their assessment briefs were limited more to women's health and mental health trauma respectively. Neither of those reviews concluded that there was much advantage in extending their brief to the crosscultural literature generally as a way of identifying assessment tools that had cultural comparability and statistical rigour for application with persons from refugee backgrounds. Both reviews and the current review concluded that construct and linguistic comparability are frequent shortcomings in efforts to develop or adapt instruments for use with persons from refugee backgrounds. In addition to giving serious consideration to those adult measures that have been shown by the present review to lend themselves to cultural and linguistic adaptation, readers are cautioned about trends in the current literature to (a) short-circuit procedures for demonstrating construct and linguistic equivalence in test development and adaptation, (b) overlook the assessment needs of child clients and (c) disregard the neuropsychological, applied cognitive and psychoeducational assessment needs of these psychologically, socially and politically vulnerable persons. Finally, valid and reliable assessment of refugees' mental health and wellness is important not only for determining need for, and appropriate choice of, mental health interventions; it also indirectly enhances clients' educational, occupational and social functioning, leading potentially to better use of services and systems and to financial sustainability. Service providers may reasonably consider assessment best practices as the first step in assisting people from refugee backgrounds to adjust personally, socially and economically to geographical and cultural transition and resettlement.

Acknowledgments

This review was prepared as background for part of an Australian Psychological Society (APS) position on refugees with assistance from the APS, but the review represents the views and summary opinions of the authors and does not constitute a formally endorsed APS position. The second author was supported by the Fulbright Association and National Cancer Institute Grants #U54 CA132384 and #U54 CA132379. The third author would like to thank the Australian Research Council for its support of a research study on the conceptualisation and treatment of people from a refugee background (ARC grant: LP0776558).

References

- Bean, T., Derluyn, I., Eurelings-Bontekoe, E., Broekaert, E., & Spinhoven, P. (2006). Validation of the multiple language versions of the Reactions of Adolescents to Traumatic Stress Questionnaire. *Journal of Traumatic Stress, 19*, 241–255.
- Bean, T., Derluyn, I., Eurelings-Bontekoe, E., Broekaert, E., & Spinhoven, P. (2007). Validation of the multiple language versions of the Hopkins Symptom Checklist-37 for refugee adolescents. *Adolescence, 42*, 51–71.
- Bean, T., Mooijaart, A., Eurelings-Bontekoe, E., & Spinhoven, P. (2007). Validation of the Teacher's Report Form for teachers of unaccompanied refugee minors. *Journal of Psychoeducational Assessment, 25*, 53–68.
- Bean, T., Mooijaart, A., Eurelings-Bontekoe, E., & Spinhoven, P. (2006). Validation of the child behavior checklist for guardians of unaccompanied refugee minors. *Children and Youth Services Review, 28*, 867–887.
- Berry, J.W. (1969). On cross-cultural comparability. *International Journal of Psychology, 4*, 119–128.
- Bolton, P. (2001a). Local perceptions of mental health effects of the Rwandan genocide. *Journal of Nervous and Mental Disease, 189*, 243–248.
- Bolton, P. (2001b). Cross-cultural validity and reliability testing of a standard psychiatric assessment instrument without a gold standard. *Journal of Nervous and Mental Disease, 189*, 238–242.
- Bolton, P., Bass, J., Betancourt, T., Speelman, L., Onyango, G., Clougherty, K.F., Neugebauer, R., Murray, L., & Verdelli, H. (2007). Interventions for depression symptoms among adolescent survivors of war and displacement in Northern Uganda: A randomized controlled trial. *Journal of the American Medical Association, 298*, 519–527.
- Bracken, P.J., Giller, J.E., & Summerfield, D. (1995). Psychological responses to war and atrocity: the limitations of current concepts. *Social Science and Medicine, 40*, 1073–1082.
- Brislin, R.W., Lonner, W.J., & Thorndike, P. (1973). *Cross-cultural research methods*. New York, NY: Wiley.
- Brough, M., Gorman, D., Ramirez, E., & Westoby, P. (2003). Young refugees talk about well-being: A qualitative analysis of refugee youth mental health from three states. *Australian Journal of Social Issues, 38*, 193–208.
- Burgoyne, U., & Hull, O. (2007). *Classroom management strategies to address the needs of Sudanese refugee learners*. Adelaide, South Australia: NCVER.
- Cha, E.S., Kim, K.H., & Erlen, J.A. (2007). Translation of scales in cross-cultural research: Issues and techniques. *Journal of Advanced Nursing, 58*, 386–395.
- Charney, M.E., & Keane, T.M. (2007). Psychometric analysis of the clinician-administered PTSD scale (CAPS)–Bosnian translation. *Cultural Diversity and Ethnic Minority Psychology, 13*, 161–168.
- Courtois, C.A. (2004). Complex trauma, complex reactions: Assessment and treatment. *Psychotherapy: Theory, Research, Practice, and Training, 41*, 412–425.
- Dana, R.H. (2005). *Multicultural assessment: Principles, applications, and examples*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Dana, R.H. (2007). Refugee assessment practices and cultural competency training. In J.P. Wilson & C.S.-k. Tang (Eds.), *Cross-cultural assessment of psychological trauma and PTSD* (pp. 91–112). New York: Springer.
- Dao, M. (1991). Designing assessment procedures for educationally at-risk Southeast Asian-American students. *Journal of Learning Disabilities, 24*, 594–601, 629.
- Davidson, G.R. (1997). The ethical use of psychological tests: Australia. *European Journal of Psychological Assessment: ITC Bulletin, 13*, 132–139.
- Davidson, G.R., Murray, K.E., & Schweitzer, R.D. (2008). Review of refugee mental health and wellbeing: Australian perspectives. *Australian Psychologist, 43*, 1601–74.
- Drożdek, B., & Wilson, J.P. (2007). Wrestling with the ghosts from the past in exile: Assessing trauma in asylum seekers. In J.P. Wilson & C.S.-k. Tang (Eds.), *Cross-cultural assessment of psychological trauma and PTSD* (pp. 113–131). New York: Springer.
- Ecklund, K., & Johnson, W.B. (2007). Toward cultural competence in child intake assessments. *Professional Psychology: Research and Practice, 38*, 356–362.
- Ehnholt, K.A., & Yule, W. (2006). Practitioner review: Assessment and treatment of refugee children and adolescents who have experienced war-related trauma. *Journal of Child Psychology and Psychiatry, 47*, 1197–1210.
- Fawzi, M.C.S., Pham, T., Lin, L., Nguyen, T.V., Ngo, D., Murphy, E., & Mollica, R.E. (1997). The validity of post-traumatic stress disorder among Vietnamese refugees. *Journal of Traumatic Stress, 10*, 101–108.
- Fraine, N., & McDade, R. (2009). Reducing bias in psychometric assessment of culturally and linguistically diverse students from refugee backgrounds in Australian schools: A process approach. *Australian Psychologist, 44*, 16–26.
- Gagnon, A.J., Tuck, J., & Barkun, L. (2004). A systematic review of questionnaires measuring the health of resettling refugee women. *Health Care for Women International, 25*, 111–149.
- Herlihy, J., Scragg, P., & Turner, S. (2002). Discrepancies in autobiographical memories—implications for the assessment of asylum seekers: repeated interviews study. *British Medical Journal, 324*(7333), 324–327.
- Hinton, D.E., Chhean, D., Pich, V., Safren, S.A., Hofmann, S.G., & Pollack, M.H. (2005). A randomized controlled trial of cognitive-behavior therapy for Cambodian refugees with

- treatment-resistant PTSD and panic attacks: A cross-over design. *Journal of Traumatic Stress*, 18, 617–629.
- Hinton, D.E., Pich, V., Safren, S.A., Pollack, M.H., & McNally, R.J. (2006). Anxiety sensitivity among Cambodian refugees with panic disorder: A factor analytic investigation. *Anxiety Disorders*, 20, 281–295.
- Hoffmann, C., McFarland, B.H., Kinzie, J.D., Bresler, L., Rakhlin, D., Wolf, S., & Kovas, A.E. (2005). Psychometric properties of a Russian version of the SF-12 Health Survey in a refugee population. *Comprehensive Psychiatry*, 46, 390–397.
- Hollifield, M., Eckert, V., Warner, T.D., Jenkins, J., Krakow, B., Ruiz, J., & Westermeyer, J. (2005). Development of an inventory for measuring war-related events in refugees. *Comprehensive Psychiatry*, 46, 67–80.
- Hollifield, M., Warner, T., Lian, N., Krakow, B., Jenkins, J.H., & Kesler, J. (2002). Measuring trauma and health status in refugees: A critical review. *Journal of the American Medical Association*, 288, 611–621.
- Kagee, A., & Naidoo, A.V. (2004). Reconceptualizing the sequelae of political torture: Limitations of a psychiatric paradigm. *Transcultural Psychiatry*, 41, 46–61.
- Kirmayer, L. J. (2003). Failures of imagination: the refugee's narrative in psychiatry. *Anthropology and Medicine*, 10, 167–185.
- Kivling-Boden, G., & Sundbom, E. (2003). Cognitive abilities related to post-traumatic symptoms among refugees from the former Yugoslavia in psychiatric treatment. *Nordic Journal of Psychiatry*, 57, 191–198.
- Lavik, N.J., Laake, P., Hauff, E., & Solberg, Ø. (1999). The use of self-reports in psychiatric studies of traumatized refugees: Validation and analysis of HSCL-25. *Nordic Journal of Psychiatry*, 53, 17–20.
- Malekzai, A.S.B., Niazi, J.M., Paige, S.R., J Hendricks, S.E., Fitzpatrick, D., Leuschen, M.P., & Millirnet, C. R. (1996). Modification of CAPS-1 for diagnosis of PTSD in Afghan refugees. *Journal of Traumatic Stress*, 9, 891–898.
- Mallinckrodt, B., & Chia-Chih, W. (2004). Quantitative methods for verifying semantic equivalence of translated research instruments: A Chinese version of the Experiences in Close Relationships Scale. *Journal of Counseling Psychology*, 51, 368–379.
- Miller, K., Omidian, P., Quraishy, A. S., Quraishy, N., Nasiry, M.N., Nasiry, S. et al. (2006). The Afghan Symptom Checklist: A culturally grounded approach to mental health assessment in a conflict zone. *American Journal of Orthopsychiatry*, 76, 423–433.
- Mollica, R.F., Caspi-Yavin, Y., Bolline, P., Truong, T., Tor, S., & Lavelle, J. (1992). Validating a cross-cultural instrument for measuring torture, trauma, and posttraumatic stress disorder in Indochinese refugees. *Journal of Nervous and Mental Disease*, 180, 111–116.
- Muecke, M.A. (1992). New paradigms for refugee health problems. *Social Science and Medicine*, 35, 515–523.
- Mulder, P.L. (1992). An investigation of the effects of level of acculturation on the performance of adult Hmong refugees on Raven's Standard Progressive Matrices and Bender-Gestalt Visual/Motor Test. *Dissertation Abstracts International*, 52(9-B), 4981.
- Odejide, A.O. (2006). Status of drug use/abuse in Africa: A review. *International Journal of Mental Health Addiction*, 4, 87–102.
- Okawa, J.B. (2008). Considerations for the cross-cultural evaluation of refugees and asylum seekers. In L.A. Suzuki & J.G. Ponterotto (Eds.), *Handbook of multicultural assessment: Clinical, psychological and educational applications* (3rd ed.; pp. 165–194). San Francisco: Jossey-Bass.
- Papadopoulos, R.K. (2007). Refugees, trauma and adversity-activated development. *European Journal of Psychotherapy and Counselling*, 9, 301–312.
- Ponterotto, J.G. (1996). Evaluating and selecting research instruments. In F.T. Leong & J.T. Austin (Eds.), *The psychology research handbook* (pp. 73–84). Thousand Oaks, CA: Sage.
- Renner, W., Salem, I., & Ottomeyer, K. (2006). Cross-cultural validation of measures of traumatic symptoms in groups of asylum seekers from Chechnya, Afghanistan, and West Africa. *Social Behavior and Personality*, 34, 1101–1114.
- Robertson, C.L., Halcon, L.L., Savik, K., Johnson, D., Spring, M., Butcher, J. et al. (2006). Somali and Oromo refugee women: Trauma and associated factors. *Journal of Advanced Nursing*, 56, 577–587.
- Rousseau, C., Crepeau, F., Foxen, P., & Houle, F. (2002). The complexity of determining refugeehood: A multidisciplinary analysis of the decision-making process of the Canadian Immigration and Refugee Board. *Journal of Refugee Studies*, 15, 43–70.
- Rousseau, C., Drapeau, A., & Corin, E. (1996). School performance and emotional problems in refugee children. *American Journal of Orthopsychiatry*, 66, 239–251.
- Sack, W.H., Seeley, J.R., Himb, C., & Clarke, G.N. (1998). Psychometric properties of the Impact of Events Scale in traumatized Cambodian refugee youth. *Personality and Individual Differences*, 25, 57–67.
- Sechrest, L., Fay, T., & Zaida, S. (1972). Problems of translation in cross-cultural research. *Journal of Cross-Cultural Psychology*, 3, 41–56.
- Silove, D. (2006). The impact of mass psychological trauma on psychosocial adaptation among refugees. In G. Reyes & G.A. Jacobs (Eds.), *Handbook of international disaster psychology: Refugee mental health* (Vol. 3, pp. 1–17). Westport, CT: Praeger Publishers.
- Sinaiko, H.W., & Brislin, R.W. (1973). Evaluating language translations: Experiments on three assessment methods. *Journal of Applied Psychology*, 57, 328–334.
- Smith, P., Perrin, S., Dyregrov, A., & Yule, W. (2003). Principal components analysis of the Impact of Events Scale with children in war. *Personality and Individual Differences*, 34, 315–322.
- Summerfield, D. (1999). A critique of seven assumptions behind psychological trauma programmes in war-affected areas. *Social Science and Medicine*, 48, 1449–1462.
- Suzuki, L.A., & Ponterotto, J.G. (Eds.). (2008). *Handbook of multicultural assessment: Clinical, psychological and educational applications* (3rd ed.). San Francisco: Jossey-Bass.
- Suzuki, L.A., Ponterotto, J.G., & Meller, P.J. (Eds.). (2001). *Handbook of multicultural assessment: Clinical, psychologi-*

- cal and educational applications* (2nd ed.). San Francisco: Jossey-Bass.
- van de Vijver, F.J.R., & Poortinga, Y.H. (2005). Conceptual and methodological issues in adapting tests. In R. K. Hambleton, P.F. Merenda, & C.D. Spielberger (Eds.), *Adapting educational and psychological tests for cross-cultural assessment* (pp. 39–63). Mahwah, NJ: Erlbaum.
- Victorian Department of Education and Early Childhood Development. (2009). *Guidelines for psychologists: Intellectual disability. Appendix 1: Assessment of refugees and recent arrivals from non-English speaking backgrounds*. Melbourne, Australia: Author.
- Weinstein, C.S., Fucetola, R., & Mollica, R. (2001). Neuropsychological issues in the assessment of refugees and victims of mass violence. *Neuropsychology Review*, 11, 131–141.
- Willis, G., & Gonzalez, A. (1998). Methodological issues in the use of survey questionnaires to assess the health effects of torture. *Journal of Nervous and Mental Disease*, 186, 283–289.
- Wilson, J.P., & Tang, C.S.-k. (Eds.). (2007). *Cross-cultural assessment of psychological trauma and PTSD*. New York: Springer.
-